Statistics and Application {ISSN 2454-7395 (online)}
Volume 20, No. 1, 2022 (New Series), pp 75-87

Lexicon Based Sentiment Analysis (LBSA) to Improve the Accuracy of Acronyms, Emoticons, and Contextual Words

T. Nikil Prakash and A. Aloysius
Department of Computer Science, St. Joseph's College, Thiruchirappalli.

Received: 02 August 2020; Revised: 30 January 2021; Accepted: 08 February 2021

Abstract

Sentiment analysis (SA) is a process of text analysis and is also called an area of Natural Language Processing (NLP). NLP is concerned with finding the emotions or opinions within the text. NLP is used to classify the opinions with the user's expression in text reviews and to analyze whether the user text is positive or negative. The data are drawn from various resources such as web blogs, social media sites, e-commerce, online banking, etc. The research work focuses on the lexicon method and the data are collected from Twitter comments data using the Kaggle website. A new Senti_Con_Acro algorithm is proposed in this paper. The proposed work is focused on sentiment acronyms, emoticons, and contextual sentiment. The result of the proposed work is to improve the accuracy and efficiency of existing work.

Key words: Sentiment analysis; Lexicon based approaches; Contextual words; Emoticons; Acronyms.

1. Introduction

Sentiment analysis (or opinion mining) is a natural language processing technique used to determine whether the data is positive, negative, or neutral. Sentiment analysis on textual information is also done to allow companies to track consumer feedback on brand and product sentiment and to understand customer needs. The advancement of social media and an increasingly wide range of communicating networks are sharing ideas and opinions among people everywhere in the world. Such comments and reviews are believed to essential assets for the users (Sankar, 2019). In recent years, a large number of people are accessing the Internet and social media. Smart devices have empowered users to share their expressions or opinions explicitly over social media and this information can reach a large audience in a fraction of seconds.

Sentiment analysis relate to the identification and other information on the feelings and attitudes expressed in natural language texts, opinions, and beneficial evaluations. It is well established and subject to considerable research to identify information relating to products, companies, and other commercial entities automatically. Sentiment analysis can be useful for handling customer feedback automatically, for concentrating advertising and for analyzing consumer trends and trends (Mullen, 2006).

Sentiment Analysis provides tools to analyse this information such as user feelings, emotions, product reviews, social media chats, comments and posts. SA has become popular in online communities to find business giants in mining consumer minds and enhanced business

Corresponding Author: T. Nikil Prakash
Email: nikilprakash1992@gmail.com

performance in recent years. It is considered to be an effective method for classifying user reviews into either positive or negative polarities. Data analytics researches and organizations to find the opinion of customers thinking and emotions. Sentiment analysis tools have identified their feeling and emotions to text. Mostly sentiment tools have been done at sentence and document level reviews. SA has three main approaches namely lexicon-based method, Machine Learning (ML) and hybrid approaches Liu (2015).

The lexicon-based method is tagged with polarity detection and the word or phrases whether determine positive or negative. ML tools are used to classify the training dataset. It automatically identifies the product reviews, online banking, etc (Devika, 2016). The hybrid approach of sentiment analysis exploits both Machine Learning methods and lexicon-based methods for polarity detection (Katrekar 2010).

Many challenges are similar to those of traditional feeling analysis, but they are not always identical. It is well known that people have ambiguous expressions of their sentiments and opinions. The Proposed work will identify the contextual words, acronyms, and emoticons in the text. The Literature Review describes existing work and a comparison of existing works and the proposed work describes the Senti_Con_Acro algorithm model and the experiments results. Finally, the conclusion and references are the final sections in this research work.

2. Literature Review

Yousif (2017) proposed and analyzed scientific citation sentiment analysis challenges and issues. They classified citation function and citation recommendation to have huge consideration of Sentiments. They identified data preprocessing methods used in scientific SA and presents citation context extraction, future extraction, and user's data sources.

Xie (2017) proposed maximum entropy probabilistic latent semantic analysis (ME-PLSA). This method extracts seed emotion words from Wikipedia and training corpus data. The maximum entropy model tests the process of emotion classification. *K*-fold model divides the training set and the test set. The emotional classification method classifies words such as relevant of words, part of speech in context, some similarity emotional words, and the relevance of degree adverbs and so on.

Felipe (2018) proposed a recommendation process in SA to textual data. They classified Facebook and Twitter datasets and classified all negative review posts. They evaluated the issues of data scarcity in e-commerce. The recommended system increased the assertiveness of the recommendation process. Support Vector Machine (SVM) algorithm classified the dataset and increased the performance of real data.

Chen (2018), proposed pre-trained character embedding with a Dual-Channel Convolutional Neural Network (char-DCCNN) in Sentiment analysis. They represent vector-matrix using the input text in a two-channel convolutional neural network. One channel is static and another one is fine-tuned. They collected the microblog sentiment dataset and identified the reviews such as film, sports, social and other filed of datasets. The Char-DCCNN method classified this dataset.

Dasa (2018) proposed real-time SA of Twitter streaming data for stock prediction. Streaming data find the source of data analysis collected in real-time. Streaming data normally deals with a continuous flow of data. The data carries information such as websites, social media, mobile phone applications, server logs, *etc*. The active learning algorithm analyzed the data and predicted user behavior in a ceaseless manner. Long short-term memory helps to stream online data prediction and provided better results.

Thelwall (2017) proposed a lexicon sentiment analysis for identifying gender biases. The lexical sentiment algorithm detects strong positive from female categories and negative from male categories. A social media monitoring algorithm is used to find the male and female attitudes in social media. They found out that the people's opinions and the SentiStrength tool gave a better performance. They compared both male and female different levels of communication in sentiment analysis and accurate data.

Hassan (2016) proposed contextual semantic sentiment analysis for Twitter data. They detected the sentiment analysis of two levels such as entity level and tweet-level. They described three data sets such as StentiStrength during lexical words and the term of strength that is fixed data or unchanged data and another one is SentiCircles it dynamically changes the contextual words. They used SentiCircles data sets for lexicon-based sentiment identification at both levels of sentiment detection. These approaches enhanced the performance of other datasets.

Alexander (2010) proposed sentiment classification using an automatic corpus collection method in the training set. They used TreeTagger for POS tagging and compared the difference in polarity sets. They used synthetic structures, it is described to emotions and or opinion state facts. They collected corpus training data sets and classified the data as positive, negative, and neutral. The classification method such as naïve Bayes uses the N-gram and Part of Speech (POS) tagging method.

Raghavendra (2019) proposed the Rule-Based Modeling (RBM) method for sentiment lexicon analysis. They collected from the dataset of Cornell's review data and identified the data whether it is positive or negative reviews in text files. The RBM classified the data and gave better performance results in existing lexicon methods. The RBM is highly regarded as both a sentiment analysis and feature extraction tools.

Table 1: Comparative details of existing work

Authors	Approaches/ Classification	Dataset	Advantages	Disadvantages
Jiangfeng, (2019)	• Aspect-Level Sentiment Classification, • NLP, • Gaussian kernel, • Information retrieval	Social Media	• The sentence is hidden on the LSTM layer. • Improved the performance of aspect-level sentiment classification	• Influence among different aspects when one opinionated sentence owns more than one aspect terms
Reinald, (2018)	• Sentence classification. • NLP	Social Media	• Classification performance and achieves state-of-the-art performance	• Complex NLP tasks. • Inaccurate translations thus data producing

				noisy sentence vector
Al-Kabi (2018)	• Arabic text classification • Opinion mining • Naïve Bayes • SVM • K-NN	Arabic	• Effectiveness of two free online tools such as social mention and SentiStrength	• Unable to extract the stem of all slang words • Dataset is not stemmed • The problem of spelling mistakes and repetition of letters and characters
Parinda (2019)	• Machine learning techniques • Information Extraction • NLP • Naive Bayes • POS tagging	Social Media	• Flexible and customizable way of generating connections between data sources	Low quality data size
Douglas, (2013).	• Dictionary-based approaches • Conjunction-based approach	Blog posts	• Applied to older corpuses • Develop a class of minimally-supervised	• The size of the dictionary • The sensitivity of various dictionaries
Jonathan, (2017),	• Machine Learning • Semantic parsing	Social media	• Parse utterances in unseen domains by decoupling structure mapping • Improved generalization such as dependency trees, syntactic CCG parses	• Structure distribution in the target domain is very different from the source • Datasets where only denotations are provided • Average accuracy
Hassan, (2016),	• Supportvector machine(SVM) • Maximumentropy(maxent)	Twitter	• Higher performance in detecting neutral entities	• Influence on performance. • Different sentiment orientations
Mike, (2017).	Phrase-level sentiment analysis	Twitter	• Automatically Identify the contextual polarity for a large subset	Low accuracy

3. Proposed Work

The Proposed work section highlights on analysis of Twitter data, based on lexicon approaches. The Senti_Con_Acro algorithm framework model is proposed in this research work. Figure 1. shows the proposed Senti_Con_Acro algorithm framework model. The proposed Senti_Con_Acro algorithm model process contains several phases that follow as:

3.1. Senti_Con_Acro model framework

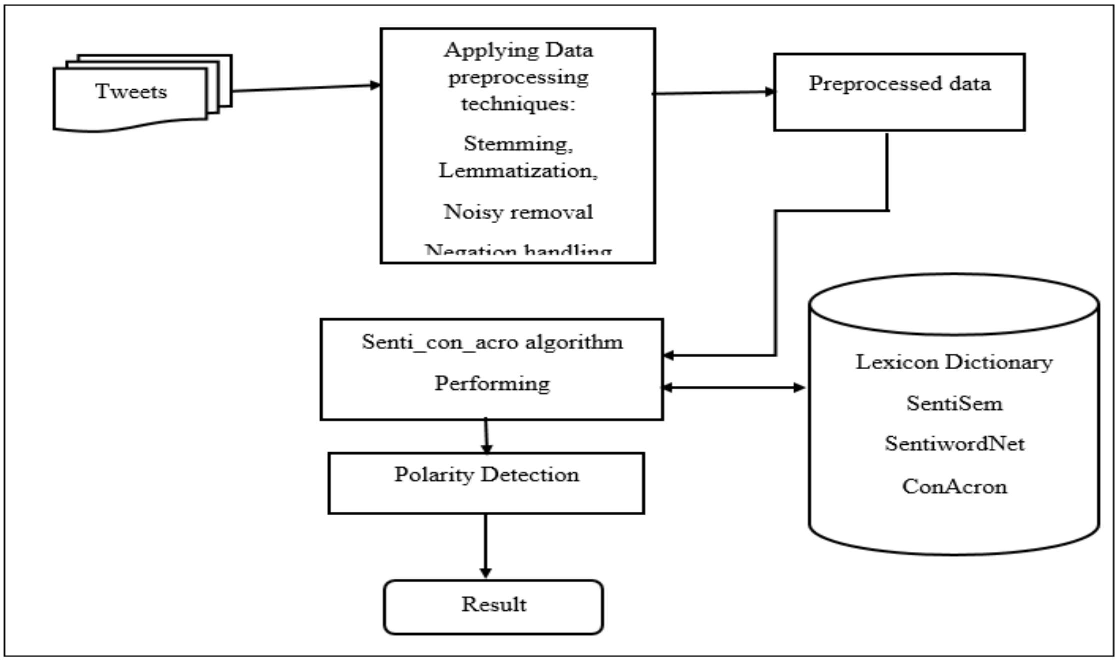

Figure 1: Framework for Senti_Con_Acro algorithm

Phase 1: Data Preprocessing

Data Preprocessing is a process of making unstructured data into structured data. It is often inconsistent, incomplete, and contains a lot of certain behaviors or product reviews and data many errors. Data preprocessing contains data preparation and data cleaning.

Data Preparation

Data preparation is a collection of a dataset from various resources like social media, e-commerce, web blogs, *etc*. In the proposed work, the data is collected from the Twitter dataset which is collected from the Kaggle website.

Data Cleaning

Data cleaning is the next step of data preprocessing. It refers to the process of cleaning to missing values, noisy data and inconsistent data. It also removes stop words and non-English words, *etc*. the Twitter dataset is cleaned and processed from the proposed work. The data cleaning contains three important steps that are as follows:

Tokenization

Tokenization is the process of splitting longer text into small pieces of text. It is also called text segmentation or lexical analysis.

Stemming

Stemming is a process of reducing words to the word stem from the dataset. Stemming algorithms have two types namely Porter stemmer and Lancaster Stemmer. Porter stemming removes morphological words and Lancaster stemmer removes aggressive words. It is used to determine domain vocabularies in domain analysis.

Lemmatization

Lemmatization is to access lexical knowledge bases and to get the correct sentence from the words. It is used for the WorldNet corpus and stops words in the corpus.

Phase 2: Feature Extraction

Feature extraction is the second stage of the proposed work. The text features are extracted from different ways that are as follows:

i. Unigram feature

In the proposed work Unigram features assume that the occurrence of each word is independent of its previous word. The proposed work counts all the input words using the unigram method. Hence each word becomes a gram (feature) here. For example:
"I", "have", "a", "lovely", "dog."

ii. Contextual words

The contextual words are called the different set of words or phrases. The proposed work identifies the user behavior or product performance in sentiment analysis. The contextual dictionary increasing content constantly which provides unmatched opportunities to support decision-making processes and advocacy efforts. Table 2 shows some examples of contextual words to convert acronyms of dictionary words.

Table 2: Contextual words to a dictionary word

Context Words	Dictionary Words
Abrupt	Sudden
Up-To-Date	Informed
Percepts	Perceived
Present	Existing
Common	Public
Constant	Continual

iii. Emoticons

The emoticons are emojis that identify the user behaviors and expressions. There is n number of emojis/ emoticons available in the emoticon dictionary. These emojis identify the positive and negative expressions in sentiment analysis algorithms. The proposed work identifies the emoticons and determines them as a positive or negative value. Figure 2 displays sentiment polarity using emoticons.

Figure 2: Emoticon Dictionary

iv. Acronyms

Abbreviations or acronyms are widely used in text materials to reduce space. The text in such areas consists of one to two sentences or a few sentences such as text messages, social media comments and blog posts. Customers may use or add new abbreviations or short word types, i.e. fast communication acronyms which rarely appear in regular or modern text, for these messages. Text as "TIA" for "Thank You in advance" is, for instance, common in these fields and for the machine. The textual significance of the texts could hardly be accurately understood. The high-rate text adds new abbreviations that can impact the reliability of the emotional analysis. To solve this problem, abbreviations must be extracted and identified before the sentiment method is performed.

Phase 3: Polarity Detection

Polarity detection is the third stage of the proposed work. Sentiment polarity determines three types of sentiment analysis that is positive, negative, and neutral. Sentiment polarity returns the overall opinion of a text or document in one single issue. The opinions classify into two opposing sentiment polarities are called positive or negative or introduce as neutral while the position of opinion locates between these two polarities. An opinionated text and categorizing it based on overall positive, neutral, and negative classes is called sentiment polarity classification.

Phase 4: Frequency Occurrence

Once the feature is extracted, they are used as input for supervised lexicon-based approaches for further classification. Generally, the frequency of occurrence of a keyword is a more suitable feature in overall sentiment analysis and not necessarily indicated by repeated use of keywords.

Brevity's law (also called Zipf's law) states that if words of a language are sorted in the order of decreasing frequency of usage, a word's frequency is inversely proportional to its rank, or sequence number in the list. The Brevity's Mandelbrot law equation is federated as the frequency of the sentiments which is measured as low rank and high-rank ratio and categorized through the deviancy of the power law. Brevity's Mandelbrot law check the ranking value if k >k0 is greater than the k0 value which gives the ranking is same in order the k value less than k0 the value is added as k0 + k

(DT) $\leftarrow f_k \infty(k0 + k)\text{-}b$
where, f ← frequency of a word,
k ← ranking of a word
DT ← Input data

Algorithm: Senti_Con_Acro

Phase 1: Pre-Processing

```
Input:Tweets (DT)
Output: Processed Tweets <dt'1, dt'2, .., dt'n> equivalent to <dt1,dt2,dt3…dtn>
Begin
        for each tweet do
                apply unigram and skip-gram // feature selection
                remove URLs, username.
                replace a repeated character by two or more indication of the same
        character
        if w is a stop word next
                remove w from DT
        else if w is a neighboring word next
                Detect the contextual word from the DT to assign W
        end if
        end for
return processed Tweets DT'
End
```

Phase 2: Feature Extraction

```
If DT word found in the dictionary then
        If DT word is a Neighboring word then
                Replace the equivalent word in a contextual word
        Else if DT word is an acronym then
                Replace the equivalent word for acronyms
        Else if DT word is emoticons then
                Replace the equivalent word for emoticons
        Else
         Identify the word is acronyms, and emoticons insert into the dictionary with
        equivalent meaning.
Else
        Identify the word is acronyms, and emoticons insert into the dictionary with
equivalent meaning.
```

Phase 3: Polarity Detection

For each DT word
DT word polarity ← DTp +DTne+ DTnu
DTp ← $\sum_{i=1}^{n} dt\, p(i)$
DTne ←$\sum_{i=1}^{n} dt\, ne(i)$
DTnu ← $\sum_{i=1}^{n} dt\, nu(i)$
$\sum_{i=1}^{n} dt$ Polarity ← sum ($\sum_{i=0}^{n} dt\, p(i)$, $\sum_{i=0}^{n} dt\, ne(i)$, $\sum_{i=0}^{n} dt\, nu(i)$)
For each dt
If dt p(i) > dt nu(k) > DT Class (i) ← positive
Else if dt ne(i) < dt nu(i) > DT Class (i) ← Negative
Else dt nu(i) ← Neutral
End for
If DTp > DTne && DT nu then DT is positive impact
Else if DTne< DTnu and DTp then DT is negative impact
Else
DT is Neutral
End for

Phase 4: Frequency Occurrence

For each DT type
Frequency occurrences (DT) ←$f_k \infty (k + k)$-b
Max(f) then
Rank ← min
End for
End

3.2. Results

3.2.1. Confusion matrix

A confusion matrix is a table that is often used to describe the performance of a classification model on a set of test data for which the true values are known. Table 3. Shows the classification of a test dataset produces four outcomes – true positive, false positive, true negative, and false negative.

Table 3: Confusion matrix classification of a test dataset

	Predicted Positives	Predicted Negatives
Actual Positive Instances	Number of True Positive	Number of False Negatives
Actual Negative Instances	Number of False Positive	Number of True Negatives

- **True Positives (TP)** - These are the correctly predicted positive values which mean that the value of the actual class is yes and the value of the predicted class is also yes.

- **True Negatives (TN)** - These are the correctly predicted negative values which mean that the value of the actual class is no and value of the predicted class is also no.

- **False Positives (FP)** – When the actual class is no and the predicted class is yes.

- **False Negatives (FN)** – When actual class is yes but predicted class is no.

3.2.2. Basic measures derived from the confusion matrix

Various measures can be derived from a confusion matrix. These are follow as:

- **Precision** - Precision is the ratio of correctly predicted positive observations to the total predicted positive observations. High precision relates to the low false-positive rate. In this proposed work done 0.7875 pretty good precision.

Precision = TP/TP+FP

- **Recall (Sensitivity)** - Recall is the ratio of correctly predicted positive observations to all observations in actual class - yes. In this proposed work is done recall of 0.7552 is good for this model as it's above 0.5.

Recall = TP/TP+FN

- **F-Measure** – F-measure is the weighted average of Precision and Recall. Therefore, this score takes both false positives and false negatives into account. If the cost of false positives and false negatives are very different, it's better to look at both Precision and Recall. In the proposed work done the F-measure is 0.7717.

F1 Score = 2*(Recall * Precision) / (Recall + Precision)

- **Accuracy** - Accuracy is the most intuitive performance measure and it is simply a ratio of correctly predicted observation to the total observations. For the proposed model, done 0.8093 which means the proposed model is approx. 81% accurate.

Accuracy = TP+TN/TP+FP+FN+TN

3.2.3. Comparison results of the existing work

Table 4: Comparison results of the proposed work and existing work

S. No	Author	Accuracy (%)
1	Seydeh Akram Saadat Neshan (2020)	76.3
2	Ahmad A. Al-Oqaily (2020)	68
3	M. Edison (2017)	68.75
4	Vallikannu Ramanathan (2019)	76
5	Proposed Work	80.93

3.2.4. Result for the proposed work

Table 5: Result for the Proposed Work

	Precision (%)	Recall (%)	F-Measure (%)	Accuracy (%)
Proposed Result	78.75	75.52	77.17	80.53

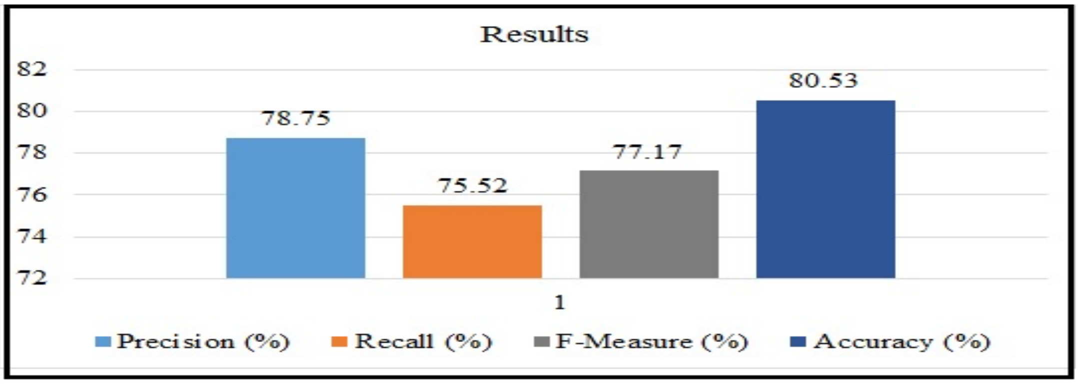

Figure 3: Evaluation measures from the confusion matrix

In this proposed work the statistical analysis of confusion matrix is applied to predict the result of emoticons, acronyms, contextual words with acronyms, and sentiments. Mainly sentiment analysis results are also predicted. The results are compared with the existing work and it brings better results than existing work.

4. Conclusion

Sentiment analysis is a platform of text analysis. It's identifying people's opinions, emotions, and sentiments, etc. This paper focused on lexicon-based sentiment analysis. A new Senti_Con_Acro Algorithm has been proposed to identify sentiment acronyms, emoticons, and contextual words. The emoticons, acronyms, contextual acronyms, and sentiments have been evaluated. This proposed work has given better results than the existing work. In the future, an image-based emotion detection method using different sentiment analysis approaches can be carried out to find the sentiments and to improve the accuracy to handle different evaluation metrics.

References

Akriti, S., Madhurima, H., Saru, D. and Madhulika, B. (2018). *An Initiative to Identify Depression using Sentiment Analysis: A Machine Learning Approach*. Indian Journal of Science and Technology, DOI: 10.17485/ijst/2018/v11i4/119594, **11(4)**, 1-6.

Alexander, P. and Patrick, P. (2010). *Twitter as a Corpus for Sentiment Analysis and Opinion Mining*. Proceedings of the International Conference on Language Resources and Evaluation, 17-23.

Al-Kabi, M., Izzat, A., Rawan, K. and Heider, W. (2018). *Evaluating Social Context in Arabic Opinion.* The International Arab Journal of Information Technology, **15**, 1-7.

Al-Oqaily, A., Al-Hassan, M., Salah, K. and Elshqeirat, B. (2020). *Sentiment Analysis for Arabic Tweets Datasets: Lexicon-based and Machine Learning Approaches*. Journal of Theoretical and Applied Information Technology, **98**(**4**), 612-624.

Ashish, K. (2010). *An Introduction to Sentiment Analysis*. Global Logic Inc. 1-6.

Bing, L. (2015). *Sentiment Analysis.* Cambridge University Press, DOI: https://doi.org/10.1017/CBO9781139084789, 1-384.

Devika, M. D., Sunitha, C. and Amal, Ganesh (2016). *Sentiment Analysis: A Comparative Study on Different Approaches.* Fourth International Conference on Recent Trends in Computer Science and Engineering, DOI:10.1016/j.procs.2016.05.124. **87(2016)**. 44-49.

Douglas, R. and Christopher, Z. (2013). *Corpus-based Dictionaries for Sentiment Analysis of Specialized Vocabularies*. Prepared for presentation at the New Directions in Analyzing Text as Data Workshop, 1-17.

Edison, M. and Aloysius, A. (2017). Lexicon based Acronyms and Emoticons Classification of Sentiment Analysis (SA) on Big Data. *International Journal of Database Theory and Application*, **10**(**7**), 41-54. DOI: http://dx.doi.org/10.14257/ijdta.2017.10.7.04,

Felipe, G., Contratres, Solange, N., Alves S., Lucia, Vilela, Leite, F. Luiz, S. and Souza, D. (2018). *Sentiment Analysis of Social Network Data for Cold-Start Relief in Recommender Systems*. Springer International Publishing AG, DOI: https://doi.org/10.1007/978-3-319-77712-2_12, 122–132.

Hassan, S., Yulan H., Miriam F. and Harith A. (2016). *Contextual Semantics for Sentiment Analysis of Twitter.* Elsevier. DOI: http://dx.doi.org/10.1016/j.ipm.2015.01.0050306-4573, **52**, Issue 1, 5-19.

Jiangfeng, Z., Xiao, M. and Zhou, K. (2019). *Enhancing Attention-Based LSTM with Position Context for Aspect-Level Sentiment Classification.* IEEE, **7**, 20462 – 20472.

Jonathan, H., Michal, Shmueli, S. and Konopnicki, D. (2017*). Emotion Detection from Text via Ensemble Classification Using Word Embedding's.* Conference: the ACM SIGIR International Conference, DOI: 10.1145/3121050.3121093, 269-272.

Mike, T. (2017). *Gender Bias in Sentiment Analysis. Emerald Publishing, Information Review,* DOI: https://doi.org/10.1108/OIR-05-2017-0139, **42**, Issue. 1, 45-57.

Parinda, P. and Premkumar, Sivakumar. (2019). *Context Dependency Relation Extraction Using Modified Evolutionary Algorithm Based on Web Mining*. Springer Nature Singapore. DOI: https://doi.org/10.1007/978-981-13-1498-8_23, 259-267.

Raghavendra, K. D., Mukul, A., Surendra, K., Keshari and Ashwini, K. (2019). *Sentiment Analysis and Feature Extraction Using Rule-Based Model (RBM).* Springer Nature Singapore, DOI: https://doi.org/10.1007/978-981-13-2354-6_7, 57-64.

Reinald, K., Amplayoy, Kyungjae, Leey, Jinyeong, Yeoz and Seung-won, Hwangy (2018). *Translations as Additional Contexts for Sentence Classification.* Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence (IJCAI-18), 1-7.

Sankar, H., Subramaniyaswamy, V., Vijayakumar, V., Sangaiah Arun Kumar, R., Logesh and Umamakeswari, A. (2019). *Intelligent Sentiment Analysis Approach using Edge Computing-based Deep Learning Technique.* Wiley Online Library, DOI: https://doi.org/10.1002/spe.2687, 1–13.

Seydeh, A., Saadat, N. and Reza, A. (2020). *A Combination of Machine Learning and Lexicon Based Techniques for Sentiment Analysis.* 6th International Conference on Web Research (ICWR), IEEE, DOI: 10.1109/ICWR49608.2020.9122298, 8-14.

Shuai, W., Mianwei, Z., Sahisnu, M., Bing, L. and Chang, Y. (2018). *Disentangling Aspect and OpinionWords in Target-based Sentiment Analysis using Lifelong Learning.* Research Gate, DOI: arXiv: 1802.05818, 1-7.

Sisi, C., Youdong, D., Zhifeng, X., Shenkai, L. and Huiming, D. (2018). *Chinese Weibo Sentiment Analysis Based on Character Embedding with Dual-Channel Convolutional Neural Network*. IEEE International Conference on Cloud Computing and Big Data Analysis, DOI: 978-1-5386-4301-3/18/$31.00, 107-111.

Suman, S. and Rajitha, N. (2018) *Unsupervised Deep Representations for Learning Audience Facial Behaviors*. IEEE, DOI: arXiv:1805.04136, 1-6.

Sushree, D., Ranjan, K., Behera, Mukesh, K., Santanu, K. and Ratha. (2018). *Real-Time Sentiment Analysis of Twitter Streaming data for Stock Prediction*. Science Direct, DOI: 10.1016/j.procs.2018.05.111, 956-964.

Tony. M. and Robert, M. (2006). *A Preliminary Investigation into Sentiment Analysis of Informal Political Discourse.* Computational Approaches to Analyzing Weblogs, Papers from the 2006 AAAI Spring Symposium, 1-4.

Vallikannu, R., Meyyappan, T. (2019). *Twitter Text Mining for Sentiment Analysis on People's Feedback about Oman Tourism*. 4th MEC International Conference on Big Data and Smart City (ICBDSC), IEEE, 1-5.

Xie, X., Songlin, G., Hu, F., Xie, M. and Jiang, N. (2017). *An Improved Algorithm for Sentiment Analysis based on Maximum Entropy.* Soft Computing, Springer. DOI: 10.1007/s00500-017-2904-0, 1-13.

Yousif, A., Niu, Z., Taurus, J. K. and Ahmad, A. (2017). *A Survey on Sentiment Analysis of Scientific Citations.* Artificial Intelligence Review. Springer, DOI: https://doi.org/10.1007/s10462-017-9597-8, 1-34.